دسته‌بندی توده‌های آب در خلیج چابهار با استفاده از روش خوشه‌بندی

بشرا وفایی[۱*]، وحید چگینی[۲]، عباس سقایی[۳]، مجتبی عظام[۴]

۱- کارشناسی ارشد فیزیک دریا، دانشکده علوم و فنون دریایی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، استان تهران، تهران، پست الکترونیکی: boshra_vafaie@yahoo.com

۲- دکترای مهندسی سواحل، استادیار، مرکز ملی اقیانوس‌شناسی، استان تهران، تهران، پست الکترونیکی: v_chegini@inco.ac.ir

۳- دکترای مهندسی صنایع، دانشیار دانشکده مهندسی صنایع، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، استان تهران، تهران، پست الکترونیکی: a_saghaei@srbiau.ac.ir

۴- دکترای فیزیک دریا، استادیار دانشکده علوم و فنون دریایی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، استان تهران، تهران، پست الکترونیکی: ezam@srbiau.ac.ir

تاریخ دریافت: ۹۰/۴/۱ ** نویسنده مسوول* *تاریخ پذیرش: ۹۱/۳/۳۰*

© نشریه علمی ـ پژوهشی اقیانوس‌شناسی ۱۳۹۲، تمامی حقوق این اثر متعلق به نشریه اقیانوس‌شناسی است.

چکیده

امروزه با گسترش پایگاه‌های ثبت داده‌های دریایی و حجم بالای داده‌های ذخیره شده، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش کرد. وقتی که حجم داده‌ها بالا باشد، نمی‌توان الگوهای مفید را در میان حجم انبوه داده‌ها تشخیص داد. بنابراین نیاز به روش‌های داده‌کاوی (Neural Network & Data Mining) است که اصطلاحا به کشف دانش بپردازند و با کمترین دخالت کاربر و به صورت خودکار الگوهای مفید در داده‌ها و رابطه‌های منطقی را پیدا نمایند. یکی از رایج‌ترین روش‌های داده‌کاوی در فیزیک دریا، خوشه‌بندی (Clustering) است. جهت خوشه‌بندی خلیج چابهار بر اساس چگالی از الگوریتم دو مرحله‌ای (Two-step) استفاده شد. همچنین برای بررسی تاثیرات دما و شوری در توزیع مکانی و زمانی چگالی آب خلیج و یافتن مؤثرترین پارامتر بر چگالی آب خلیج از میان پارامترهای موثر (عمق، دما و شوری)، یازده شبکه عصبی در نرم افزار کلمنتاین (Clementine) تشکیل شد. نتایج به‌طور واضحی نشان دهنده‌ی یک مورد استثنا در آبان ماه است که با رسم منحنی‌های همتراز چگالی، شوری و دما توسط نرم‌افزار سرفر (Surfer) نتایج به‌دست آمده از داده‌کاوی تایید شد.

کلمات کلیدی: *چابهار، چگالی، داده‌کاوی، خوشه‌بندی، شبکه عصبی، کلمنتاین، سرفر.*

۱. مقدمه

با توجه به اهمیت راهبردی مرزهای دریایی و حوضه‌های آبی وهمچنین نیاز روز افزون به ثبت اطلاعات دریایی در این مناطق به نظر می‌رسد امروزه جامعه دریایی کشور با پایگاه‌هایی حجیم از داده‌های دریایی روبرو است و اطلاعات حاصل از پردازش این داده‌ها جهت بهره‌برداری بهینه از مناطق دریایی صرف‌نظر از جنبه‌های زیست محیطی و شیلاتی، از نظر توسعه حمل و نقل دریایی ، طراحی بهینه سازه‌های دریایی، احداث بنادر صیادی،

تجاری، نظامی و احداث سکوهای نفتی و انجام دیگر فعالیت‌های مهندسی حائز اهمیت است.

بنابراین با گسترش پایگاه‌های داده‌های دریایی و حجم بالای داده‌های ذخیره شده، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش کرد و اطلاعات حاصل از این پردازش را در اختیار کاربرانی نظیر موسسه ملی اقیانوس شناسی، موسسه تحقیقات شیلات، سازمان محیط زیست و سازمان بنادر و دریانوردی قرار داد. همان‌طور که در کشورهای پیشرفته سال‌ها است که روش‌های نوین داده‌کاوی توجه اقیانوس‌شناسان و مهندسین دریایی را به خود جلب کرده است، به‌نظر می‌رسد که وقت آن رسیده که با استفاده از روش‌های نوین داده‌کاوی به کشف دانش از میان داده‌ها پرداخت و با کمترین دخالت کاربر و به‌صورت خودکار، الگوهای مفید در بین داده‌ها و رابطه‌های منطقی را یافت. برای مثال به‌منظور شناخت خصوصیات و پارامترهای فیزیکی و شیمیایی آب و بررسی تغییرات فصلی این خصوصیات، تعیین الگوی توزیع این پارامترها و شناسایی لایه‌های آب و مورد پراهمیت دیگر یعنی پیش‌بینی رفتار پارامترهای فیزیکی و شیمیایی آب، از این روش‌ها استفاده نمود. طبیعت متغیر پارامترهای فیزیکی آب، بیان آماری آن‌ها را ایجاب می‌کند، لذا دوره اندازه‌گیری اطلاعات فیزیکی آب باید به اندازه‌ی کافی طولانی باشد که زمینه برای تحلیل‌های آماری و احتمالاتی فراهم آید. البته در مواردی که انجام این کار امکان‌پذیر نباشد نیز داده‌کاوی به کمک اقیانوس‌شناسان و مهندسان می‌آید و اطلاعات شبیه‌سازی شده به‌وسیله‌ی رویکردهای نوین داده‌کاوی به‌عنوان جایگزین ثبت‌های واقعی، به‌کار می‌رود.

برای مثال، داشتن اطلاعات CTD و جریان‌سنج[1] مسیر بهینه‌ای برای کشتی‌ها به‌منظور کمترین میزان مصرف انرژی پیش‌بینی شد (Palenzuela et al., 2010). در تحقیق دیگری برای بررسی توزیع آب دریا، روش‌های خوشه‌بندی براساس چگالی پیشنهاد گردید (Kisilevich et al., 2010). پریرا و همکاران با استفاده از سری‌های زمانی ۱۰ ساله پارامترهای فیزیکی شیمیایی و زیست‌شناختی و به‌کارگیری روش قواعد انجمنی[2]، فراجوشی آب در منطقه آرایال دو کابو(در منطق ساحلی برزیل) را تحلیل کردند (Pereira et al., 2008). دسته‌بندی توده‌های آب و لایه پراکندگی صوتی در آب‌های شمال شرق تایوان با استفاده از روش خوشه‌بندی فازی[3] در سال ۱۹۹۸ انجام شد (Liao, 1998). در دو تحقیق دیگر، با استفاده از دمای سطحی آب و ارتفاع موج شاخص و به‌کارگیری خوشه‌بندی براساس چگالی، به موضوع توزیع فضایی زمانی پارامترهای فیزیکی آب و توده‌های مشابه آب در دریای مجاور ترکیه پرداخته شد (Birant and Kut, 2006). داده کاوی فضایی در سامانه‌ی اطلاعات مکانی دریایی نیز به‌کار برده شده است (Gaixiao et al., 2010).

با استفاده از رگرسیون لجستیک که یکی از روش‌های داده کاوی است، به پیش‌بینی فراوانی نوعی ماهی دریایی در آب‌های اقیانوسی اطراف نیوزلند مبادرت شده است (Leathwick et al., 2006). همچنین با داشتن داده‌های سنجش از دور و از طریق خوشه‌بندی می‌توان چرخندها را تشخیص و مسیر حرکت آن‌ها را تعیین نمود (Stolorz et al., 1995).

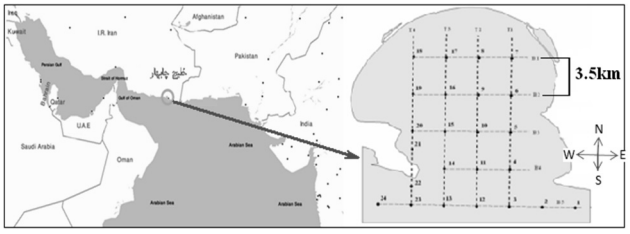
۲. منطقه‌ی مورد مطالعه

منطقه‌ی مورد مطالعه در این تحقیق، خلیج چابهار است. خلیج چابهار با طول جغرافیایی بین ″۱۵ ′۲۵ °۶۰ تا ″۴۵ ′۳۲ °۶۰ و عرض جغرافیایی ″۱۵ ′۱۷ °۲۵ تا ″۲۰ ′۲۷ °۲۵ در جنوب شرقی سواحل استان سیستان و بلوچستان قرار دارد. وسعت این خلیج بالغ بر ۳۲۰ کیلومتر مربع، عمق متوسط آن ۶ متر و عمق بیشینه‌ی آن در دهانه‌ی ورودی ۲۲ متر است. طول این خلیج در راستای شمالی جنوبی ۱۷ کیلومتر و عرض آن در راستای شرقی غربی در عریض‌ترین قسمت، حدود ۲۰ کیلومتر است و هیچ رودخانه قابل ملاحظه‌ای وارد این خلیج نمی‌شود. خلیج چابهار بزرگ‌ترین خلیج ایران در حاشیه کرانه‌های دریای عمان و به خاطر شکل امگا شکل (Ω) خود در زمین‌شناسی از نوع خلیج‌های امگایی نامیده می‌شود.

قرار گرفتن این بندر در مسیر بادهای موسمی شبه قاره هند موسوم به مونسون در تابستان چابهار را به خنک‌ترین و در زمستان آن را به گرم‌ترین بندر جنوبی کشور تبدیل کرده است. میانگین درجه حرارت بندر چابهار برابر با $۲۷^{°C}$ است. درجه‌ی رطوبت فقط در دو ماه از سال (اردیبهشت و خرداد) در این مکان بالا می‌رود. میانگین بارندگی سالیانه در این شهرستان برابر با ۱۰۵

[1] Current meter
[2] Association rules
[3] Fuzzy Clustering Method

میلی‌متر ثبت شده که ۶۴٪ آن مربوط به ماه‌های دی و بهمن و اسفند است. همچنین حداکثر رطوبت ثبت شده ۶۷٪ و حداقل آن ۵۴٪ است (کمیجانی، ۱۳۸۷). شکل ۱ نمایانگر موقعیت این خلیج در دریای عمان است.



شکل ۱: موقعیت خلیج چابهار در دریای عمان (مقیاس تصویر سمت راست ۱:۷۳۷۷۰۵)

۲-۱. انتخاب ایستگاه‌های نمونه‌برداری

در بررسی و مطالعه منابع آبی، یکی از مهم‌ترین مراحل تحقیق، تعیین ایستگاه‌های نمونه‌برداری است. به‌طور کلی ملاک انتخاب ایستگاه‌های نمونه‌برداری با توجه به اهداف عملیات به‌گونه‌ای طراحی شده که ضمن پوشش قرار دادن بخش وسیعی از آب‌های خلیج، اعماق مختلف را نیز در برگرفته و وسعت بیشتری از آب‌های منطقه مورد مطالعه را شامل است. بدین منظور تعداد ۴ خط نمونه‌برداری عمود بر ساحل[۱] و هر کدام به فاصله ۲ دقیقه‌ای از یکدیگر (حدود ۳/۵ کیلومتر) انتخاب و بر روی هر ترانسکت ۵ ایستگاه به فواصل ۲ دقیقه‌ای تعیین گردید. همچنین به‌منظور تحلیل بهتر، خلیج به سه نوار شمالی (نوار ۱)، مرکزی (نوار ۲) و جنوبی (نوار ۳) تقسیم شد. در مجموع ۱۷ ایستگاه داخل خلیج با فواصل ۲ دقیقه‌ای و ۶ ایستگاه با شماره‌های ۱، ۲، ۳، ۱۲، ۱۳،۲۳ خارج از خلیج تعیین گردیدند. جانمایی ایستگاه‌های انتخاب شده در شکل ۱ نشان داده شده است.

۲-۲. روش اندازه‌گیری پارامترها و روش برداشت داده‌های CTD

طی سال ۱۳۸۷ دستگاه CTD مدل Ocean seven 316 ساخت شرکت ایدرونات کشور ایتالیا جهت بررسی خصوصیات فیزیکی آب دریا در خلیج چابهار در زمان‌های از پیش تعیین شده در ایستگاه‌های مورد نظر، با سرعت یک متر در ثانیه به عمق آب فرستاده و با گام زمانی یک ثانیه داده‌برداری انجام شد. این عمل در کل سطح خلیج برای ۲۳ ایستگاه انجام گردید.

بدین منظور پارامترهای عمق متناسب با فشار، دما، هدایت الکتریکی برداشت شد که به‌وسیله‌ی آنها شوری (psu) و چگالی (kg/m^3) نیز محاسبه شدند. از آن‌جایی که نمونه‌برداری‌ها در زمان‌های متفاوتی و با تراز جزر و مدی مختلف انجام شد، عمق نمونه‌برداری از ایستگاه‌های تعیین شده در طی این۱۱ دوره متفاوت بوده است.

۲-۳. معرفی داده‌کاوی

امروزه با گسترش سامانه‌های پایگاهی و حجم بالای داده‌های ذخیره شده در این سامانه‌ها، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش کرد. وقتی که حجم داده‌ها بالا باشد، کاربران هر چند زبر دست و با تجربه باشند، نمی‌توانند الگوهای مفید را در میان حجم انبوه داده‌ها تشخیص دهند (حقیقی و همکاران، ۱۳۸۵). از سوی دیگر کاربران معمولا فرضیه‌ای را مطرح می‌کنند و سپس بر اساس گزارشات مشاهده شده به اثبات یا رد آن می‌پردازند، در حالی که امروزه نیاز به روش‌هایی است که با آن بتوان به کشف دانش پرداخت، یعنی با کمترین دخالت کاربر و به‌صورت خودکار الگوها و رابطه‌های منطقی بیان گردند. به‌وسیله داده‌کاوی که از آغاز دهه ۹۰ پا به عرصه ظهور گذاشت و از سال ۱۹۹۵ به‌صورت جدی وارد مباحث آماری شد، الگوهای مفیدی در داده‌ها با حداقل دخالت کاربران شناخته می‌شوند (Fayyad, 1996).

برای داده‌کاوی تعاریف مختلفی وجود دارد که معمول‌ترین آن عبارت از استخراج دانش و اطلاعات از یک پایگاه داده بسیار بزرگ و پیچیده (تبدیل داده‌ها به دانش) و همچنین کشف الگوهای پنهان بین داده‌ها است (Berry and Lindoff, 1997).

بنابراین ابزارهای داده‌کاوی الگوهای پنهانی را کشف و پیش‌بینی می‌کنند که متخصصان ممکن است به‌دلیل اینکه این اطلاعات و الگوها خارج از انتظار آن‌ها باشد، آن‌ها را مدنظر قرار ندهند و به آن‌ها دست نیابند (Pendharkar, 2003).

[۱] خطوط یا ردیفهای نمونه برداری عمود بر ساحل Transect

۲-۴. روش‌های داده‌کاوی برای کشف دانش

از میان روش‌های متفاوت داده‌کاوی نظیر دسته‌بندی[1] و رگرسیون[2] و تحلیل سری‌های زمانی[3] خوشه‌بندی و کشف قوانین انجمنی (وابستگی) و الگوهای متوالی[4] که در مطالعات دریایی کاربرد دارند، در این تحقیق از روش خوشه‌بندی به‌منظور تعیین ایستگاه‌های مشابه در خلیج چابهار استفاده شده است. همچنین به‌منظور تعیین موثرترین پارامتر از میان پارامترهای دما، شوری و عمق از الگوریتم شبکه عصبی استفاده شده است.

اغلب اطلاعات و داده‌های موجود در پایگاه‌های داده‌ها توزیع‌های ناشناخته یا پیچیده‌ای دارند که به‌راحتی نمی‌توان آن توزیع‌ها را شناسایی نمود. بنابراین استفاده از روش‌هایی که نیاز به دانستن توزیع متغیرها ندارد از اهمیت خاصی برخوردار است. خوشه‌بندی یکی از روشهایی است که با توزیع داده‌های موجود سر و کار نداشته و اغلب با استفاده از معیارهای تشابه و عدم تشابه به خوشه‌بندی داده‌ها می‌پردازد (مهریزی، ۱۳۸۲ و ناظمی، ۱۳۸۳). تمرکز اصلی تحلیل خوشه‌ای عمدتاً بر اساس فاصله است، حساسیتی به‌ترتیب وارد شدن داده‌های ورودی ندارد. همچنین نتایج خوشه‌بندی قابل تفسیر و قابل فهم است (حقیقی و همکاران، ۱۳۸۵).

خوشه‌بندی در واقع یک عملیات غیر نظارتی[5] است. این عملیات هنگامی استفاده می‌شود که هدف به‌دست آوردن گروه‌هایی از داده‌های مشابه باشد، بدون اینکه از قبل یک پیش‌بینی در مورد شباهت‌ها موجود باشد. کاربرد خوشه‌بندی یافتن گروه‌هایی از روابط بین متغیرها است که قبلا شناخته نشده‌اند. در خوشه‌بندی شباهت رکوردهای هر خوشه بسیار زیاد است و به‌عبارت دیگر فاصله رکوردها از همدیگر کم است و برعکس شباهت خوشه های مجاور کم است (فاصله خوشه های مجاور از همدیگر زیاد است) (Hasti et al., 2008).

اگرچه در خوشه‌بندی معمولا از الگوریتم‌های K-means یا K-medoids و غیره استفاده می‌کنند، اما روش مورد نظر از هر نوعی که باشد، عملیات انجام شده یکسان است (ایوزیان، ۱۳۸۳).

در این تحقیق از خوشه‌بندی با الگوریتم دو مرحله‌ای[6] استفاده شده است.

این الگوریتم یک تحلیل خوشه‌ای قابل مقیاس‌بندی است که برای میزان داده زیاد ابداع شده است. با این روش می‌توان متغیرهای پیوسته و categorical یا attributes را بررسی نمود.

همان‌گونه که از نام این روش مشخص است این الگوریتم شامل دو مرحله است:

۱- پیش‌خوشه‌بندی[7] کردن حالت‌ها (یا اطلاعات) به تعدادی زیرخوشه[8]

این گام از یک روش خوشه‌بندی متوالی استفاده می‌کند. آن داده‌ها را یک به یک بررسی نموده و تصمیم می‌گیرد که آیا داده تحت بررسی می‌تواند در دسته‌بندی‌های قبلی قرار گیرد یا یک دسته‌بندی جدید را براساس معیار فاصله که در زیر توضیح داده شده، آغاز کند.

۲- خوشه‌بندی کردن زیرخوشه‌های به‌دست آمده از مرحله اول. تعداد خوشه‌ها به‌صورت خودکار در این مرحله تعیین می‌شود. زمانی یک داده در حال بررسی به خوشه‌ای اختصاص داده می‌شود که به آن نزدیک‌تر است. سپس کل خوشه‌ها مقایسه شده و هر جفت خوشه با کمترین فاصله مابینشان انتخاب شده و در یک خوشه ادغام می‌شوند.

در این تحقیق برای طراحی تمام شبکه های عصبی از روش[9] RBFN استفاده شده است. روش مذکور داده‌ها را بر اساس فیلد هدف که در این تحقیق چگالی در نظر گرفته شده است افراز می‌کند. رویکرد آن مشابه روش k-means در خوشه بندی است.

از آنجایی که هدف شبکه عصبی دسته بندی اطلاعات ورودی است، بنابراین یک روش نظارتی است که به ازای هر دسته از الگوهای ورودی خروجی‌های متناظر نیز به شبکه نشان داده می‌شود و تغییر وزن‌ها تا موقعی صورت می‌گیرد که اختلاف خروجی شبکه به ازای الگوهای آموزشی از خروجی‌های مطلوب در حد خطای قابل قبولی باشد. در این روش برای تخمین و اصلاح وزن‌ها، از روش حداقل مربعات خطی استفاده می شود و هدف طرح شبکه‌ای است که ابتدا با استفاده از داده‌های آموزشی موجود، آموزش ببیند و سپس با ارائه بردار ورودی به شبکه،

[1] Classification
[2] Regression
[3] Time Series Analysis
[4] Sequential Patterns
[5] Unsupervised Operation
[6] Two-Step Cluster Algorithms
[7] pre-cluster
[8] sub-cluster
[9] Radial Basis Function Network

کلاس آن را تشخیص دهد. چنین شبکه‌ای به‌طور گسترده برای کارهای تشخیص الگو به‌کار گرفته می‌شود. تابع بکار رفته از نوع سیگموییدی[1] است.

۳. روش تحقیق

در این تحقیق ضمن معرفی داده‌کاوی و ارایه چندین مورد از کاربردهای آن در بحث‌های دریایی، دو مورد از روش‌های داده‌کاوی نظیر خوشه‌بندی و شبکه عصبی برای داده‌های سال ۱۳۸۷ استفاده شده و سعی شده تا به یکی ازطریق‌های زیر اقدام کرد:

۱- با خوشه‌بندی خلیج از نظر چگالی:

الف- ایستگاه‌های مشابه را در ماه‌های مختلف دسته‌بندی نمود.

ب- برای شناسایی لایه‌های آب و تعیین خصوصیات فیزیکی آب از قبیل دما، شوری و چگالی در اعماق مختلف در سرتاسر خلیج چابهار در کل سال ۱۳۸۷ اقدام نمود که توسط این روش تعداد خوشه‌ها به‌صورت خودکار چهار خوشه تعیین شده است.

۲- همچنین به‌منظور بررسی اهمیت متغیرها و یافتن مهم‌ترین عامل تاثیرگذار از میان پارامترهای دما و شوری بر روی چگالی آب و تعین میزان تأثیرآن پارامتر از روش بررسی اهمیت متغیر توسط الگوریتم شبکه‌ی عصبی که یک روش نظارتی در نرم‌افزار کلمنتاین (از زیر مجموعه‌های SPSS) است استفاده گردید. برای این منظور در هر ماه به‌صورت جداگانه اثر دما و شوری بر روی چگالی بررسی شد تا با لحاظ کردن عامل زمانی نتایج دقیق‌تری به‌دست آید.

از آن‌جایی که در زمان تشکیل شبکه، چگالی به‌عنوان پارامتر هدف[2] انتخاب شده بود، در نتیجه درجه اهمیت هر یک از متغیرهای دما و شوری که بر توزیع چگالی تاثیر گذارند برحسب درصدی از عدد یک به‌دست آمد. نتایج (جدول ۱) به‌طور واضح نشان دهنده‌ی این موضوع است که در اکثر ماه‌ها موثرترین پارامتر بر نحوه توزیع چگالی خلیج، دما است و شوری از درجه‌ی اهمیت کمتری نسبت به دما برخوردار است. البته یک مورد استثنا نیز در آبان ماه وجود دارد.

۳- برای بررسی اعتبار و صحت این روش، آزمون آماری رگرسیون (روش ریج)[3] به‌کار برده شد. بدین منظور با توجه به کمیت‌های برداشت شده رگرسیون گیری در کل خلیج و برای هر ماه به‌صورت جداگانه انجام گرفت که در این روابط، عمق، درجه حرارت و شوری پارامتر مستقل و چگالی پارامتر وابسته در نظرگرفته شده است. در نتیجه در هر دوره یک رابطه‌ی ریاضی به‌دست آمد. نتایج به‌دست آمده نشان داد که می‌توان از شبکه عصبی پاسخهای قابل قبولی دریافت نمود.[4]

۴- به‌دلیل آن‌که رسم خطوط هم‌مقادیر در لایه‌های عمقی مورد نظر، شناخت دقیق‌تری از سیمای منطقه‌ی مورد مطالعه و مقایسه پارامترها در لایه‌های مختلف را در دسترس قرار می‌دهد، نمودارهای مربوط به خطوط هم‌تراز[5] دما، شوری، چگالی در دو لایه‌ی سطحی و بستر با استفاده از نرم‌افزار سرفر با روش درونیابی کریجینگ در منطقه ترسیم گردید. به این ترتیب که محور افقی نشان دهنده‌ی طول جغرافیایی و محور عمودی عرض جغرافیایی و منحنی‌ها نشان‌دهنده‌ی خطوط هم‌دما، هم‌شوری، هم‌چگالی و هم‌جریان هستند. اطلاعات به‌دست آمده از این مطالعه، در تعیین خصوصیات هیدرودینامیکی خلیج نظیر جریان‌ها و تعیین الگوی گردش آب مفید خواهد بود.

۴. نتایج

نتایج یازده شبکه عصبی تشکیل شده در نرم افزار Clementine در جدول ۱ آمده است.

جدول ۱: نتایج بررسی آثار دما و شوری بر چگالی براساس داده‌کاوی (Data mining)

مرحله	اهمیت دما	اهمیت شوری	دقت تخمین	لایه ورودی	لایه پنهان	لایه خروجی
۱	۰/۸۶۷۲	۰/۱۳۲۸	۹۶/۶۹۸	۲ نرون	۳ نرون	۱ نرون
۲	۰/۸۹۸۱	۰/۱۰۱۹	۹۶/۳۰۹	۲ نرون	۳ نرون	۱ نرون
۳	۰/۹۹۱۸	۰/۰۰۸۲	۹۶/۱۹۷	۲ نرون	۳ نرون	۱ نرون
۴	۰/۹۱۸۳	۰/۰۸۱۷	۹۴/۷۸۲	۲ نرون	۳ نرون	۱ نرون
۵	۰/۵۰۳	۰/۴۹۷	۹۷/۱۶۹	۲ نرون	۳ نرون	۱ نرون
۶	۰/۶۳۵۷	۰/۳۶۴۳	۹۷/۰۲	۲ نرون	۳ نرون	۱ نرون
۷	۰/۴۲۳۲	۰/۵۷۶۸	۹۴/۳۰۸	۲ نرون	۳ نرون	۱ نرون
۸	۰/۷۵۶۹	۰/۲۴۳۱	۹۴/۴۶	۲ نرون	۳ نرون	۱ نرون
۹	۰/۵۴۴۹	۰/۴۵۵۱	۹۴/۸۷۱	۲ نرون	۳ نرون	۱ نرون
۱۰	۰/۵۶۵۵	۰/۴۳۴۵	۹۳/۱۱۶	۲ نرون	۳ نرون	۱ نرون
۱۱	۰/۷۴۴۹	۰/۲۵۵۱	۹۲/۳۷۵	۲ نرون	۳ نرون	۱ نرون

با توجه به اعداد موجود در جدول ۱:

[1] Sigmoid
[2] Target
[3] Ridge method

[4] دلیل استفاده از روش ریج بجای رگرسیون ساده این است که ضرایب بدست آمده از آزمون تحلیل همبستگی پیرسون بوسیله نرم‌افزار SPSS نشان دهنده همبستگی زیاد بین پارامترهای فیزیکی آب در این خلیج می باشد.

[5] Contour

– در ماه‌های فروردین، اردیبهشت، خرداد، مرداد، آذر و اسفند دما بسیار با اهمیت‌تر از شوری است. این میزان اهمیت در ماه‌های مهر، دی و بهمن به مراتب کمتر شده و در شهریور تقریبا دما و شوری از درجه اهمیت یکسانی برخوردارند.

– این در حالی است که درجه‌ی اهمیت کمیت دما در آبان ماه از شوری حتی کمتر است. مقدار اهمیت دما از ۰/۹۱۸۳ (که درجه‌ی اهمیت بیشتر از ۹۰٪بوده است) در مرداد کاهش یافته و به ۰/۴۲۳۲ در آبان رسیده به این معنی که درجه‌ی اهمیت کمتر از ۵۰٪ شده است.

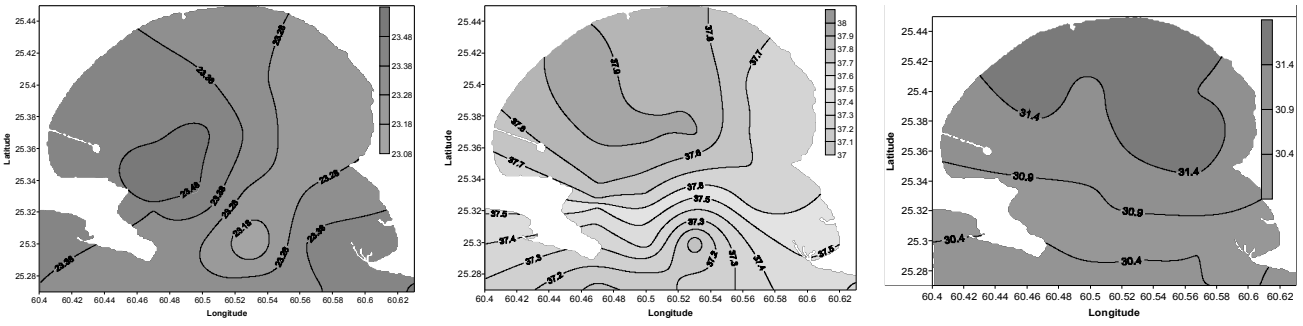
برای بررسی صحت این امر و اطمینان از بهینه بودن نتایج شبکه عصبی، منحنی‌های هم‌تراز چگالی، شوری و دما در هر ماه توسط نرم افزار سرفر رسم گردید (شکل‌های ۲ و ۳) و نتایج زیر حاصل شد:

– خطوط هم‌تراز رسم شده علاوه بر تطابق از نظر نحوه شکل‌گیری، از نظر فشردگی خطوط نیز فوق‌العاده با یکدیگر مشابهت دارند و در اکثر ماه‌ها نحوه توزیع چگالی خلیج به شدت از توزیع دما پیروی می‌کند. معمولا در محدوده‌ای که توده‌ی آب گرم و پر شور قرار دارد، توده‌ی پر چگال وجود ندارد. در حالی که در شهریور ماه ساختار تغییر کرده و در ناحیه‌ای از خلیج که دما و شوری بالایی دارد، چگالی نیز بالا رفته و این مورد در آبان ماه به اوج شدت خود می‌رسد و توده‌ای گرم، پرشور و پر چگال در نواحی ساحلی خلیج مشاهده می‌شود. تعریف خوشه‌ها در جدول ۲ آمده است.

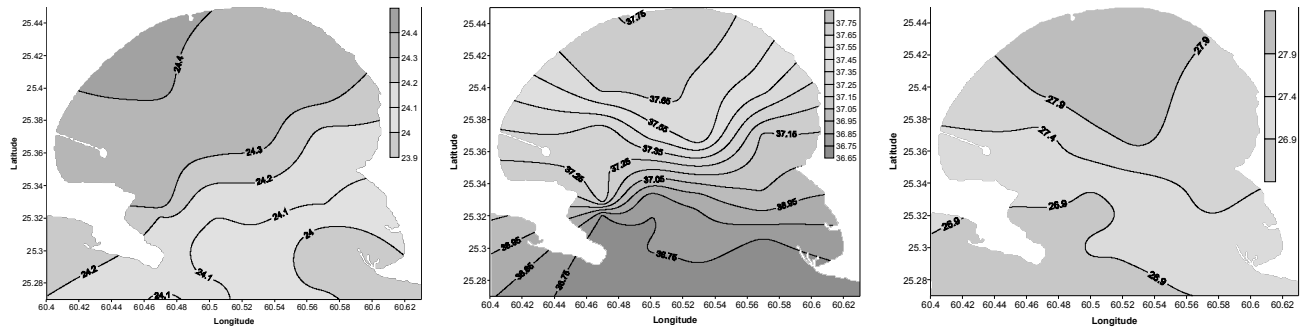
جدول ۲: تعریف خوشه‌ها

		عمق m	دما °C	شوری psu	چگالی kg/m^3
خوشه ۱	بیشینه	۱۳/۷۷	۲۵/۹۴۸	۳۷/۳۶۳	۲۵/۸۲۶
	کمینه	۰/۱	۲۱/۵۷۳	۳۶/۱۹۳	۲۴/۲۳
خوشه ۲	بیشینه	۲۴/۰۶	۳۱/۶۳۴	۳۷/۲	۲۴/۴۶
	کمینه	۰/۱۱	۲۵/۵۷۴	۳۶/۳۰۳	۲۲/۴۸
خوشه ۳	بیشینه	۲۷/۷۶	۲۸/۸۶۴	۳۷/۰۰۳	۲۵/۴۹
	کمینه	۹/۴۳	۲۱/۶۹۹	۳۶/۰۹۱	۲۳/۴۲۴
خوشه ۴	بیشینه	۱۴/۲۵	۳۲/۴۰۷	۳۸/۰۵۴	۲۴/۵۸
	کمینه	۰/۱	۲۷/۰۴۶	۳۶/۷۴۶	۲۲/۳۶

منظور از نواحی ساحلی ایستگاه‌های ۶، ۷، ۸، ۹، ۱۶، ۱۷، ۱۸ و ۱۹، نواحی مرکزی شامل ایستگاه‌های ۴، ۵، ۱۰، ۱۱، ۱۴، ۱۵، ۲۰ و ۲۱ و دهانه ایستگاه‌های ۱، ۲، ۳، ۱۲ و ۲۳ است. به‌منظور مقایسه شرق و غرب خلیج، باند شرقی شامل ایستگاه‌های ۴، ۵، ۶، ۷، ۸، ۹، ۱۰ و ۱۱ و غرب خلیج شامل ایستگاه‌های ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰ و ۲۱ است. طبق جدول ۳ مشخص است که آب‌های سطحی تر سبک‌تر بوده زیرا در خوشه‌های ۲ و ۴ اکثرا جای گرفته‌اند، در حالی‌که در آب‌های عمقی‌تر نظیر ۱۰ تا ۱۵ متر و خصوصا در ۱۵ تا ۲۷ متری آب‌های سنگین‌تر خوشه ۳ و ۱ را نیز شاهد هستیم.



شکل ۲: منحنی‌های همتراز دما، شوری و چگالی به‌ترتیب از راست به چپ در شهریور ماه



شکل ۳: منحنی‌های همتراز دما، شوری و چگالی به‌ترتیب از راست به چپ در آبان ماه

جدول ۳: نتایج خوشه‌بندی

ماه	آب سطحی (سطح تا عمق یک متر)							آب در اعماق یک تا ۵ متر							آب در اعماق ۵ تا ۱۰ متر						
	ساحل	مرکز	دهانه	شرق	غرب	ساحل	مرکز	دهانه	شرق	غرب	ساحل	مرکز	دهانه	شرق	غرب	ساحل	مرکز	دهانه	شرق	غرب
فروردین	۴و۲	۲	۲	۲	۴و۲	۴و۲	۲	۲	۲	۴و۲	۴و۲	۲	۲	۲	۴و۲	۴و۲	۲	۲	۲	۴و۲
اردیبهشت	۴و۲	۴و۲	۲	۴و۲	۴و۲	۴و۲	۴و۲	۲	۴و۲	۴و۲	۴و۲	۴و۲	۲	۴و۲	۴و۲	۴و۲	۱و۴و۲	۱و۲	۴و۲	۴و۲
خرداد	۴و۲	۴و۲	۴و۲	۴و۲	۴و۲	۴و۲	۲	۴و۲	۴و۲	۴و۲	۴و۲	۲	۲	۲	۴و۲	۴و۲	۲	۲	۲	۴و۲
مرداد	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴
شهریور	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴
مهر	۴و۲	۴و۲	۲	۴و۲	۴و۲	۴و۲	۴و۲	۱و۲	۴و۲	۴و۲	۴و۲	۴و۲	۱و۲و۳	۴	۴و۲	۴و۲	۴و۲	۴و۲	۴و۲	۴و۲
آبان	۴	۲	۲	۴و۲	۴و۲	۴	۲	۲	۴و۲	۴و۲	۴	۲	۲	۲	۴	۴	۲	۲	۴و۲	۴و۲
آذر	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
دی	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
بهمن	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
اسفند	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱

ماه	آب در اعماق ۱۰ تا ۱۵ متر							آب در اعماق ۱۵ تا ۲۷ متر			
	ساحل	مرکز	دهانه	شرق	غرب	ساحل	مرکز	دهانه	شرق	غرب
فروردین	۴و۲	۲	۳و۲	۲	۴و۲	–	۳و۲	۳	۳	–
اردیبهشت	۲	۳و۲	۳و۲	۳و۲	۳و۲	–	۳	۳	۳	–
خرداد	۴و۲	۲	۲	۲	۴و۲	–	–	۲	–	–
مرداد	۴و۲	۴و۲	۴و۲	۴و۲	۴	–	۳	۳و۲	۳	–
شهریور	۴	۴و۳و۲	۴و۳و۲	۴و۳و۲	۴و۳و۲	–	۳	۳	۳	–
مهر	۴و۲	۳و۲	۳	۴و۳و۲	۳و۲	–	۳	۳	۳	–
آبان	۴	۲	۲	۲	۲	–	–	۳	–	–
آذر	۱	۱	۳و۱	۱	۱	–	–	۳	–	–
دی	۱	۳و۱	۳و۱	۳و۱	۱	–	–	۳	–	–
بهمن	۱	۳و۱	۳و۱	۳و۱	۱	–	۳	۳	۳	–
اسفند	۱	۳و۱	۳و۱	۳و۱	۱	–	۳	۳	۳	–

برای مثال در تمامی سطوح فروردین خوشه ۲ و ۴ بوده، در حالی‌که در آب‌های عمیق (۱۵ تا ۲۷ متری) خوشه ۳ شده است. این مورد نشان می‌دهد در اکثر ماه‌ها شاره با چگالی بیشتر در بستر و شاره کم‌چگال‌تر بر روی آن قرار دارد. بنابراین شاره‌ای با چینه‌بندی چگالی پایدار وجود دارد به‌طوری که سامانه از نظر استاتیکی پایدار است.

در اردیبهشت ماه از سطح تا ۵ متری دهانه‌ی خلیج خوشه ۲ بوده و با گذشتن از ۱۰ متر شامل خوشه ۳ نیز شده است که نشان دهنده‌ی ورود آب سرد و چگال از دریای عمان و از طریق بستر به داخل خلیج چابهار است. این توده‌ی آب از سمت ترنسکت اول حوالی ایستگاه ۳ وارد خلیج شده و به‌سمت ساحل پیشروی می‌کند از طرفی از روی نمودار مقاطع عرضی رسم شده برای دما در همین ترنسکت مشخص است که آب سرد عمان از همین ناحیه وارد خلیج شده و با پیشروی به سمت ساحل گرمتر می‌شود. در نتیجه با افزایش دما، چگالی آن کمتر شده و این توده‌ی آب به‌سمت بالا حرکت کرده و به اصطلاح فراجوشی رخ می‌دهد (جهت چرخش آب پادساعتگرد است). گرادیان‌های نسبتا خوبی در اردیبهشت ماه که زمان پیش‌مونسون است قابل مشاهده است و لایه‌بندی فصلی شروع به شکل‌گیری می‌نماید (لایه پیکنوکلاین)، در حالی که در خرداد یعنی زمان به وقوع پیوستن مونسون تابستانه، تغییرات دمایی از سطح تا عمق بسیار کاهش یافته و آب تقریبا حالت یکنواختی را پیدا کرده است به عبارتی این گرادیان‌ها از بین رفته‌اند (مطابق با جدول ۲ در این ماه نواحی ساحلی، مرکزی و دهانه فقط خوشه‌های ۲ و ۴ را شامل شده اند) . این امر ناشی از وزش بادهای موسمی تابستانه و در نتیجه اختلاط کامل آب و از بین رفتن لایه‌بندی آب است. در خرداد برخلاف روند دو ماه گذشته، در بستر خوشه ۳ نداریم و همچنان خوشه۲ را می‌توان مشاهده کرد و مفهوم آن است که علاوه بر عدم وجود لایه پیکنوکلاین، بستر خلیج نسبت به دو ماه قبل کم چگال‌تر شده ولی چون هنوز هم چگالی بستر بیشتر از چگالی سطحی است، چینه‌بندی چگالی به‌صورت پایدار است.

در مرداد ماه، در ساحل از سطح تا بستر خوشه ۲ حاکم است بنابراین لایه‌بندی دمایی مشاهده نشد، در حالی‌که در باند مرکزی و دهانه یک ترموکلاین قوی تشکیل شد؛ زیرا یک تغییر از خوشه ۲ به خوشه ۳ در حوالی عمق ۱۵ متری به بعد داریم مشابه این را برای شهریور از عمق ۱۰ متری به بعد می‌توان دید (در شهریور

در کل خلیج از سطح تا ۱۰ متری فقط خوشه ۴ را داریم و از ۱۰ تا ۱۵ متری در نواحی مرکزی و دهانه خلیج خوشه ۳ نیز به آنها اضافه شده). منحنی‌های همتراز رسم شده در سرفر (شکل ۳) نیز تایید کننده این گرادیان‌های شدید، دما، شوری و چگالی مرداد و شهریور هستند. منحنی‌ها نشان دادند که در مرداد ماه ورودی آب سرد عمان از بستر دهانه به داخل خلیج حدودا تا ۶ کیلومتری ساحل در بستر پیشروی کرده و بنابراین باند مرکزی و دهانه را پوشش می‌دهد، در حالی که باند ساحلی تحت تاثیر زبانه ورودی آب سرد عمان قرار نمی‌گیرد (نواحی شرقی و دهانه‌ی خلیج از عمق ۱۵ متری تا بستر کاملا خوشه ۳ شده است). با مقایسه این اطلاعات با شرایط سطح یعنی گرمایش آب سطحی بر اثر گرم شدن هوا در مرداد ماه و همچنین ورودی آب سطحی گرم از عمان که بر اثر مونسون تابستانه رخ داده است، اختلاف دمای قابل ملاحظه‌ای برای باند مرکزی و دهانه وجود دارد و در نتیجه آن لایه‌ی ترموکلاین تشکیل می‌شود؛ در حالی که برای باند ساحلی اختلافی قابل توجهی وجود ندارد.

در مهرماه مطابق با جدول ۲ دهانه‌ی خلیج از عمق ۱ متری به بعد خوشه ۱ و ۳ را نیز شامل شده بنابراین در مهرماه عمق شروع لایه‌بندی آب، بالاتر آمده است و آب از حدود ۳ تا ۴ متری لایه‌بندی شده است (ضخامت لایه آمیخته حدود ۳ متراست). در حالی که این عمق برای ماه‌های مرداد و شهریور از ۹ تا ۱۵ متر بود است. در مهرماه ترموکلاین نسبتا قوی وجود دارد.

در آبان ماه از سطح تا ۱۵ متری برای باند ساحلی خوشه ۴ و برای باند مرکزی و دهانه خوشه ۲ را داریم بنابراین لایه‌بندی چگالی در این ماه وجود ندارد ومشخص است که ساختار از مهر به بعد تغییر کرده است، زیرا همچنان‌که آب کم چگال عمان از سطح به داخل خلیج می‌آید، خلیج را به نواحی شمالی و جنوبی تقسیم می‌کند (برعکس ماه‌های شهریور، مرداد و تا حدودی مهر که خلیج به نواحی شرقی غربی تقسیم شده بود).

در آذرماه برای تمامی سطوح خوشه ۱ را داریم بنابراین می‌توان دریافت که ستون آب به‌صورت کاملا یکنواخت درآمده است و نیز چگالی سطحی خلیج افزایش قابل توجهی یافته، به‌طوری که بیشترین چگالی در ماه قبل اکنون به مقدار کمینه چگالی سطحی در این ماه تبدیل شده است. از آذر تا اواخر زمستان از سطح تا ۱۰ متری فقط خوشه ۱ را داریم و از ۱۰ تا ۲۷ متری برای نواحی مرکزی و دهانه‌ی خوشه ۳ را نیز داریم. بنابراین در ماه‌های سرد سال که همراه با مونسون زمستانی است اکثر ایستگاه‌ها در محدوده‌ی پرچگال قرار می‌گیرند که نتیجه‌ای کاملا قابل قبول است. در این ماه‌ها شوری تا عمق ۱۵ متری زیاد و از عمق ۱۵ تا ۲۷ متری کمتر شده است که نتیجه ورود آب‌های کم شور دریای عمان به درون خلیج است.

۵. بحث و نتیجه‌گیری

در نواحی ساحلی لایه‌بندی حرارتی و چگالی وجود ندارد و به اصطلاح این کمیت‌ها در نواحی ساحلی همگن هستند که علت آن کم بودن عمق خلیج در این باند و اغتشاشات ناشی از عمق کم در محل شکست امواج و نزدیکی این نواحی به ساحل است. همچنین عامل دیگر همگن شدن لایه‌ی سطحی و عمقی در این ناحیه، نفوذ سریع انرژی گرمایی به لایه‌های زیرین و همدمایی سریع آن است. اما در ماه‌هایی که خوشه‌بندی از ۲ و۴ در برخی سطوح به ۱ و۳ رسید، برای نواحی مرکزی و دهانه‌ی خلیج، لایه‌بندی فصلی دیده شد و مشاهده گردید که این لایه‌بندی با شروع مونسون تابستانه و زمستانه دستخوش تغییرات شدیدی می‌شود. همچنین ضخامت این لایه‌ها (ترموکلاین، پیکنوکلاین و هالوکلاین) به‌دست آمد. قرارگیری خوشه‌های ۲ و۴ در سطوح فوقانی و خوشه‌های ۱ و۳ در سطوح زیرین نشان دهنده‌ی افزایش چگالی در راستای قائم با افزایش عمق است که به‌تدریج حالتی پایدار در ستون آب به‌وجود آمد که مطابق گرایش عمومی طبیعت برای رسیدن به حالت حداقل انرژی است. در این حالت شاره‌ای با چینه‌بندی چگالی پایدار وجود دارد، به‌طوری که سامانه از نظر استاتیکی پایدار است. از آن‌جایی که نتایج به‌دست آمده از داده‌کاوی کاملا مطابق با تحلیل‌های فیزیکی به‌دست آمده از رسم منحنی‌های هم‌تراز رسم شده در نرم افزار سرفر است می‌توان به‌منظور بررسی پایداری و عدم پایداری چینه‌بندی چگالی در محیط‌های آبی، وجود لایه‌های ترموکلاین، پیکنوکلاین و هالوکلاین و نیز به‌دست آوری عمق این لایه‌ها، تعیین میزان وابستگی چگالی در هر ماه به پارامترهای دما، شوری و عمق واز همه مهمتر یافتن الگوی گردش آب از داده‌کاوی استفاده نمود.

۶. سپاسگزاری

بدین‌وسیله از مساعدت‌های موسسه ملی اقیانوس‌شناسی در مورد داده‌های برداشت شده از منطقه و کمک‌های آقایان مهندس

محمد سعید سنجانی، مهندس غلامرضا ولی‌زاده و خانم مهندس حجاریان تشکر و قدردانی می‌شود.

منابع

ایوزیان م.، ۱۳۸۳. استفاده از تکنیک‌های داده‌کاوی به منظور طبقه‌بندی مشتریان. پایان‌نامه کارشناسی ارشد. دانشکده مهندسی صنایع. دانشگاه علم و صنعت ایران.

حقیقی، ع.؛ قصوری، ش.؛ برفی‌پور، ا.؛ شماخی، ع.؛ مهریزی، ع.؛ حسنی، ح.؛ یگانگی، م.، ۱۳۸۵. داده‌کاوی و کاربرد آن در کیفیت داده‌ها. فصلنامه گزیده مطالب آماری، مرکز آمار ایران، سال ۱۷، شماره ۲، صفحات ۴۷-۷۴.

کمیجانی، ف.، ۱۳۸۷. مطالعه روند تغییرات فصلی پارامترهای فیزیکی دما، شوری و چگالی در خلیج چابهار ایران. پایان‌نامه کارشناسی ارشد. دانشکده منابع طبیعی و علوم دریایی. دانشگاه تربیت مدرس.

مهریزی، ع.، ۱۳۸۲. داده‌کاوی: مفاهیم و روش‌ها و کاربردها. پایان نامه کارشناسی ارشد. دانشگاه علامه طباطبایی تهران.

ناظمی، ع.، ۱۳۸۳. رده‌بندی و داده‌کاوی. پایان نامه کارشناسی ارشد. دانشگاه فردوسی مشهد.

Berry, M.; Lindoff, G., 1997. Mastering data mining. John Wiley and Sons.

Birant, D.; Kut, A., 2006. An algorithm to discover spatial–temporal distributions of physical seawater characteristics and a case study in Turkish seas. Journal of Marine Science and Technology, Vol. 11, No. 3, 183–192.

Birant, D.; Kut, A., 2006. Cluster analysis for physical oceanographic data and oceanographic surveys in Turkish seas. Journal of Marine Research, Vol. 64, No. 5, 651-668 pp.

Fayyad; Shapiro, U.M. P.; Smyth, P.; Uthurusamy, R., 1996. Advances in Knowledge Discovery and Data Mining. Menlo Park, California: AAAI Press.

Gaixiao, L.; Rencan, P.; Yidong, Z.; Jidong, Z., 2010. Spatial Data Mining and its application in Marine Geographical Information System. International Conference on Environmental Science and Information Application Technology (ESIAT), Wuhan.

Hasti, T., 2008. The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer, 2nd ed.

Kisilevich, S.; Mansmann, F.; Nanni, M.; Rinzivillo, S., 2010. Spatio-Temporal Clustering: a Survey. Technical Report, ISTI-CNR, Italy. Submitted to Data Mining and Knowledge Discovery Handbook, Springer.

Leathwick, J.; Elith, J.; Francis, M.; Hastie, T.; Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding Newzealand: an analysis using boosted regression trees. Marine Ecology, Progress Series 77: 802–813 pp.

Liao, C. H.; Lee, K. T.; Lee, M. A.; Lu, H. J., 1998. Classification of Water Masses and Sound-Scattering Layer Biomass in the Waters off Northeastern Taiwan Using a Fuzzy Clustering Method. Department of Fisheries Science, National Taiwan Ocean University; Keelung, Taiwan, R.O.C.

Palenzuela, J. M.; Vilas, L.G.; Spyrakos, E.; Dominguez, L. R., 2010. Routing optimization using neural networks and oceanographic models from remote sensing data. 1st International Symposium on Fishing Vessel Energy Efficiency E-Fishing , Vigo, Spain.

Pendharkar, P. C., 2003. Managing Data Mining Technologies in Organizations: Techniques and Applications Idea Group Publishing.

Pereira, G.C.; Coutinho, R.; Francisco, N.; Ebecken, F., 2008. Data mining for environmental analysis and diagnostic: A case study of upwelling ecosystem of Arraial Do Cabo. Brazilian Journal of Oceanography, Vol.56, No.1, Sao Paulo.

Stolorz, P.; Nakamura, H.; Mesrobian, E.; Muntz, R.R.; Santos, J.R.; Yi, J.; Ng, K., 1995. Fast spatio-temporal data mining of large geophysical datasets. Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD'95). AAAI Press, 300–305 pp.